

-1-

### Scheduler for a Shared Channel

#### **FIELD OF THE INVENTION**

The present invention relates to a system, method and apparatus for scheduling data in a network. More specifically, the present invention relates to a system, method and apparatus for scheduling data traffic over a shared channel.

#### **BACKGROUND OF THE INVENTION**

In a network including a shared channel to deliver data traffic to multiple receiving stations from a single transmitting station, a transmitting station must determine how to allocate its downlink capacity among receiving stations. Examples of such a network include CATV-based data networks and wireless networks such as the AMOSPHERE™ system sold by the assignee of the present invention, etc. In the latter system, a base station transceiver will service a plurality of subscriber stations through an air interface that provides both shared and dedicated downlink (base station to subscriber station) channels. Because the transmission capacity of such systems is limited (typically by the available bandwidth), allocating the available capacity among the users to ensure efficient use of the transmission capacity and providing acceptable service levels can be difficult. Accordingly, such systems can benefit from appropriately scheduling transmissions over the shared link(s).

One of the simplest scheduling methods is *round-robin* scheduling. Round-robin scheduling provides each subscriber station with an equal amount of transmission time over the shared channel. While this can be an advantageous method in some circumstances, in many networks, such as those employing radio-based links, not all subscriber stations will have the same data reception rates, due to factors such as different signal-to-noise ratios (SNR). Thus, round-robin sharing does not actually provide an equal delivery of data over the shared downlink to each subscriber station. This inequality can result in dissatisfied subscribers, particularly for those subscribers using subscriber stations at the edges of the service area who have substantially lower average data rates than subscribers with subscriber stations located close to the base station. Furthermore, this inequality requires the service provider to be conservative when advertising the performance capabilities offered by the system.

-2-

Another known scheduling method, *proportionally fair sharing*, provides each subscriber station with an adjusted amount of channel capacity on the shared link, where each subscriber station receives a channel share that is adjusted by their data reception rates, so that each subscriber station receives approximately the same average amount of data. While proportional fair sharing can provide a better degree of equality between subscriber stations, it can also lead to an overall drop in total system throughput, as the base station must devote a large amount of channel capacity to service a small minority of subscriber stations with poor SNRs.

In contrast to the above two methods which focus on providing equal service to subscriber stations, some other methods focus on achieving optimal throughput across the entire system, at the expense of fairness between subscriber stations. In their article “CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users” (IEEE Communication Magazine, July 2000, pp. 70-78), Bender *et al.* demonstrate how unequal latency between subscriber stations with different data reception capabilities (i.e. – different SNRs) can increase total throughput on the downlink in the network. By providing a greater portion of channel capacity on the shared link to subscriber stations with better instantaneous SNRs, the base station can transmit more traffic overall. This method increases the total throughput of the system, potentially clearing backlogs and improving overall network performance, but also creates latency and a lower data rate for subscriber stations with poorer average SNRs. To ensure that all subscriber stations possess at least a tolerable individual data rate, the system limits the maximum permitted latency value to subscriber stations with poorer SNR ratios.

Liu *et al.* in their article “Opportunistic Transmission Scheduling with Resource-Sharing Constraints in Wireless Networks” (IEEE Journal on Selected Areas in Communication, Vol. 19, No. 10, October 2001, pp. 2053-2064) discuss the potential of improving wireless resource efficiency by exploiting variances in channel conditions, while still maintaining a level of fairness between subscriber stations. In their model, each subscriber station is allocated a fraction of transmission time. Data packets are stored in traffic queues for each subscriber station until their scheduled transmission time occurs, like the methods described above. However, the base station continuously measures the quality of the link (which can vary over time) at each

-3-

subscriber station, and transmits to the best possible subscriber station at that moment while still providing a required average data rate to each subscriber station. While opportunistic transmission scheduling can provide an increase in both total and individual throughput, it is not without its disadvantages. One disadvantage is that opportunistic scheduling can increase latency for subscriber stations as the base station waits for the opportune time to transmit to them. Another disadvantage is that this method assumes a constant level of data to be transmitted, such as a WAP session on a cellular phone. Nor does the opportunistic method contemplate intentional differences in the treatment of subscribers, such as those provided by different qualities of service (QoS), or for treatments of different types of media data. For example, a base station may service subscriber stations with different priorities, so that one subscriber using a latency-intolerant Voice-over-IP (VoIP) service receives a guaranteed service, while another subscriber is web surfing and will be provided only with a best-effort service by the base station.

10 In their article "Providing Quality of Service over a Shared Wireless Link" (IEEE Communications Magazine Feb 2001, pp. 50-54), Andrews *et al.*, describe a new scheduling algorithm, *Modified Largest Weighted Delays First* (M-LWDF) which ameliorates some of the above disadvantages. M-LWDF is similar to opportunistic scheduling in that it schedules transmissions to take advantage of SNR fluctuations. However, particular traffic flows (traveling on the downlink to their respective subscriber stations) are given preferential treatment over other traffic flows by virtue of their QoS level. For each time slot ( $t$ ), the base station calculates a value for each traffic queue, this value consisting of the product of the packet delay multiplied by the channel capacity for that subscriber station multiplied by an arbitrary value. The traffic queue with the highest derived value is scheduled. Mathematically, each time slot ( $t$ ) serves a waiting packet queue ( $j$ ) for which the function  $\gamma_j W_j(t) r_j(t)$  is maximal.  $W_j(t)$  is the current waiting time for the packets stored in queue  $j$ ,  $r_j(t)$  is the channel capacity, or data rate, for data flow  $j$ , and  $\gamma_j$  is an arbitrary value. If  $\gamma$  is the same for each packet queue, then all subscriber stations have the same QoS level. A packet queue with a higher value for  $\gamma$  thus has a higher level of service than a packet queue with a lower value for  $\gamma$ . If  $r(t)$  is the same for each packet queue, then all subscriber stations have the same data rate. A packet queue with a higher value for  $r(t)$  thus transmits at a higher average data rate than another packet queue which has the same

$\gamma$  value. While M-LWDF provides some advantage over the prior art, it also has its limitations. A key disadvantage is that M-LWDF provides no means to provide a policy of fairness between subscriber stations with different channel quality. Another disadvantage is that the M-LWDF can schedule traffic for only a single subscriber station per time slot. This creates a significant 5 amount of latency ( $x$  number of timeslots) for all other subscribers waiting for packet delivery. Furthermore, since each timeslot bears data traffic for a single subscriber station, it is possible that some of the capacity of that timeslot is wasted due to internal fragmentation.

It is therefore desired to provide a data transmission scheduler that provides QoS-differentiated service to subscriber stations while offering service providers the ability to 10 implement scheduling policies designed to provide improved sector-throughput, fairness between subscribers, or some combination of the two. It is also desirable to implement a data transmission scheduler which is opportunistic and can take advantage of variance in the data reception characteristics of subscriber stations served by the transmitter. Finally the scheduling mechanism should be flexible enough to accommodate a wide variety of computational 15 capabilities at the base station.

#### **SUMMARY OF THE INVENTION**

It is an object of the present invention to provide a novel data transmission scheduler which obviates or mitigates at least some of the above-identified disadvantages of the prior art.

According to a first aspect of the present invention, there is provided a method of 20 scheduling data for transmission from a base station over a shared channel to a plurality of subscriber stations, said method comprising:

determining a fairness factor from a range of possible fairness factors, where a first end of said range indicates a policy of scheduling data with maximum fairness between said plurality of subscriber stations and a second end of said range indicates a policy of scheduling data for maximum data traffic over said shared channel;

for each subscriber station in said plurality of subscriber stations that said base station has data to be delivered to

determining a quality of service priority value, said quality of service

-5-

priority value indicating a priority of said subscriber station relative to other subscriber stations in said plurality of subscriber stations;

determining a throughput value, said throughput value indicating the quantity of data to be moved to said subscriber station if data is scheduled to said subscriber station;

determining a total priority value, said total priority value is the sum of said quality of service priority adjusted according to said fairness factor and said throughput value conversely adjusted according to said fairness factor; and

scheduling data on a portion of said shared channel to at least one of said plurality of subscriber stations, beginning with the subscriber station with the highest total priority value.

According to another aspect of the present invention, there is provided a system for transmitting data, comprising:

a plurality of subscriber stations each having a processor, a modem, a radio and an antenna, each subscriber station operable to transmit a request for a dedicated data channel from a base station; and

a base station having a processor, a modem, a radio and an antenna, and operable to receive said requests for a dedicated data channel from said subscriber stations and to schedule data for transmission to said plurality of subscriber stations over a shared channel in accordance with a scheduling policy that varies priorities between scheduling for fairness between subscriber stations and for improved throughput to said plurality of subscriber stations.

The present invention provides a method, system and apparatus for scheduling data to a plurality of subscriber stations from a base station over a shared channel. Data destined for each subscriber station is placed into individual queues at the base station. The base station allocates a portion of the shared channel to transmit the data in each particular queue based upon a priority value it assigns to that queue. The priority value for each queue is determined by a QoS value and a throughput value, where each of these two values is adjusted by a fairness factor. The QoS value indicates whether a subscriber station has been receiving data from the base station according to an agreed-upon QoS level. The throughput value indicates the data rate that can be

-6-

achieved by the base station transmitting to that subscriber station. The fairness factor represents a scheduling policy at the base station. Scheduling policies can include scheduling data to emphasize fairness between subscriber stations at a particular QoS level, scheduling data to maximize throughput over the shared channel, and scheduling data to achieve a balance between 5 fairness and maximum throughput.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

*Embodiments of the present invention will now be described, by way of example only, with reference to the attached Figures, wherein:*

10 Figure 1 is a schematic representation of a wireless network in accordance with an embodiment of the invention;

Figure 2 is a representation of a communications link as shown in Figure 1, comprised of multiple channels;

Figure 3 is a schematic representation of the base station shown in Figure 1;

15 Figure 4 is a schematic representation of one of the subscriber stations shown in Figure 1;

Figure 5 is a schematic representation of a scheduler for a shared channel running on the base station shown in Figure 3;

Figure 6 is a flowchart showing how the broadcast downlink channel scheduler shown in Figure 5 manages the scheduling of backlogged traffic flows to the shared channel.

#### **20 DETAILED DESCRIPTION OF THE INVENTION**

Referring now to Figure 1, a wireless network for transmitting data is indicated generally at 20. Network 20 includes a radio base station 24 and a plurality of subscriber stations 28a, 28b ... 28n. In a presently preferred embodiment, radio base station 24 is connected to at least one data telecommunications network (not shown), such as a land line-based switched data network, 25 a packet network, etc., by an appropriate gateway and one or more backhauls (not shown), such as a T1, T3, E1, E3, OC3 or other suitable land line link, or can be a satellite or other radio or microwave channel link or any other link suitable for operation as a backhaul as will occur to those of skill in the art.

-7-

Base station 24 communicates with subscriber stations 28 which can be fixed, nomadic or mobile devices. The number 'n' of subscriber stations serviced by a base station 24 can vary depending upon the amount of radio bandwidth available and/or the configuration and requirements of the subscriber stations 28.

5 As illustrated in Figure 1, the geographic distribution of subscriber stations 28 with respect to base station 24 need not be symmetric nor will subscriber stations which are physically located close to one another necessarily experience the same or similar data reception rates, due to varying signal to noise ratios (SNRs) experienced at the subscriber stations 28 due to a variety of factors including the geographic environment (the presence or absence of  
10 buildings which can reflect or mask signals), the radio environment (the presence or absence of radio noise sources), etc. Thus, in most circumstances subscriber stations 28 served by a base station 24 can have significantly different SNRs and these SNRs can change over time.

As known to those of skill in the art, subscriber stations 28 can be geographically divided into different sectors 36, formed via directional antennas at base station 24 to increase the  
15 number of subscriber station 28 that can be served from a single base station location. In such a case, each sector 36 essentially acts as a different base station and base station 24 can manage the network resources in each sector 36 independent of each other sector 36.

While Figure 1 shows only one base station 24, it will further be apparent to those of skill in the art that network 20 can contain multiple, geographically distributed base stations 24,  
20 with overlapping sector 36 coverage of subscriber stations 28, and where each subscriber station 28 in an overlapping sector 36 coverage area can select which base station 24 it will be serviced by.

A communication link 32 is established in each sector 36 between base station 24 and each subscriber station 28 in the sector 36 via radio. Communication link 32a carries  
25 information to be transferred between base station 24 and subscriber station 28b, communication link 32b carries information to be transferred between base station 24 and subscriber stations 28c and 28d, etc. Communication link 32 can be implemented using a variety of multiple access techniques, including TDMA, FDMA, CDMA or hybrid systems such as GSM, etc. In a present embodiment, data transmitted over communication link 32 is transmitted using CDMA as a

multiple access technology and the data is in the form of packets, transmitted within slotted time frames, the details of which will be discussed in greater detail below.

As used herein, the terms "package", "packaged" and "packaging" refer to the overall arrangement of the transmission of the packaged data for its reception at an intended destination receiver. Packaging of data can include, without limitation, applying different levels of forward error correcting (FEC) codes (from no coding to high levels of coding and/or different coding methods), employing various levels of symbol repetition, employing different modulation schemes (4-QAM, 16-QAM, 64-QAM, etc.) and any other techniques or methods for arranging data transmission with a selection of the amount of radio (or other physical layer) resources required, the data rate and probability of transmission errors which are appropriate for the transmission. For example, data can be packaged with rate 1/4 FEC coding (each 1 data bit is transmitted in 4 bits of information) and 16-QAM modulation for transmission to a first intended receiver and packaged with rate 1/2 FEC coding and 64-QAM modulation for transmission to a second intended receiver which has a better reception-quality than the first.

Communications link 32 operates in both an uplink (from a subscriber station 28 to base station 24) and a downlink direction (from base station 24 to subscriber stations 28). The method of providing both uplink and downlink direction is not particularly limited, and in the present embodiment communications link 32 operates by frequency division duplexing (FDD). However, other methods of providing both an uplink and downlink direction, such as time division duplexing (TDD) and hybrid schemes are within the scope of the invention.

Referring now to Figure 2, in the current embodiment, communications link 32 is comprised of a plurality of channels, which in the present CDMA implementation, is achieved with orthogonal coding of link 32. In the downlink direction, base station 24 uses a shared channel, referred to as the broadcast data channel (BDCH) 38 to carry variable-rate and bursty traffic (consisting primarily of signaling and internet traffic) across a sector 36. BDCH 38 makes use of adaptive FEC and modulation to maximize downlink capacity and contains multiple packets or, more commonly, segments of packets of data for various subscriber stations 28 all time-multiplexed together into a single frame. In the present embodiment, BDCH 38 can be configured with spreading factor 4, wherein eight blocks of data can be sent within a ten

millisecond frame, spreading factor 8 wherein four blocks of data can be sent within a frame or spreading factor 16 wherein two blocks of data can be sent within a frame. Also in the present embodiment, base station 24 can support one or more BDCH 38 per sector 36 at any one time.

In the uplink direction, data traffic is carried from subscriber station 28 to base station 24 using a dedicated data channel (DDCH) 44. A separate DDCH 44 is set-up between base station 24 and each subscriber station 28 with an active communications link 32. Signaling traffic is carried from subscriber station 28 to base station 24, typically inband using DDCH 44.

Subscriber stations 28 measure their received SNR, or other metric of their ability to receive data from base station 24 and report this information back to base station 24 on a regular basis over their DDCH 44 using an upper layer signaling protocol. Subscriber stations 28 with high SNRs require less channel coding and can use higher order modulation than subscriber stations 28 with lower SNRs and thus, each block transmitted on BDCH 38 using a different block type (i.e., different packaging ofFEC type, FEC rate, modulation, etc.). Figure 3 shows an example of a base station 24 in greater detail. For the sake of clarity, base station 24 shows an example of a single sector base station. However, multi-sector base stations 24 are also within the scope of the invention. Base station 24 comprises an antenna 50, or antennas, for receiving and transmitting radio-communications over communications link 32. In turn, antenna 50 is connected to a radio 52 and a modem 54. Modem 50 is connected to a microprocessor-router assembly 56 such as a Intel Corporation Pentium processor based system using a conventional operating system such as Linux. Microprocessor-router assembly 56 is responsible for traffic scheduling of all subscriber stations 28 within its sector 36 and for radio resource management. It will be understood that assembly 56 can include multiple microprocessors, as desired and/or that the router can be provided as a separate unit, if desired. The router within microprocessor-router assembly 56 is connected to a backhaul 58 in any suitable manner, which in turn connects base station 24 to a data network (not shown).

Referring now to Figure 4, an example of a subscriber station 28 is shown in greater detail. Subscriber station 28 comprises an antenna 60, or antennas, for receiving and transmitting radio-communications over communications link 32. In turn, antenna 60 is connected to a radio 64 and a modem 68, which in turn is connected to a microprocessor-assembly 72.

-10-

Microprocessor-assembly 72 can include, for example, a StrongARM processor manufactured by Intel Corporation, that performs a variety of functions, including implementing A/D-D/A conversion, filters, encoders, decoders, data compressors, de-compressors and/or packet disassembly. Micro-processor-assembly 72 also includes buffers 74 which stores queued 5 data traffic waiting for transport up communications link 32.

As shown in Figure 4, microprocessor-assembly 72 interconnects modem 68 and a data port 76, for connecting subscriber station 28 to a data client device (not shown), such as a personal computer, personal digital assistant or the like which is operable to use data received over communications link 32. Accordingly, microprocessor-assembly 72 is operable to process 10 data between data port 76 and modem 68. Microprocessor-assembly 72 is also interconnected to at least one telephony port 80, for connecting subscriber station 28 to a telephony device (not shown) such as a telephone. In some cases, particularly in the case of a mobile subscriber station 28, the data client device can be integrated into the subscriber station 28.

Referring now to Figure 5, the logical architecture of a shared channel, such as a BDCH 15 scheduler is shown generally at 100. Scheduler 100 is responsible for assigning queued packets of data intended to be transmitted from base station 24 to subscriber stations 28 into the bitstream of BDCH 38 while maintaining any agreed-upon QoS terms for each subscriber station 28 and implementing a scheduling policy based upon a fairness factor provided by a network operator (and discussed in greater detail below) in order to provide varying degrees of 20 prioritization of fairness between subscriber stations 28 and overall throughput on BDCH 38. A method for scheduler 100 to implement a scheduling policy and schedule queued packets is described further below with reference to Figure 6. In the current embodiment, scheduler 100 is a software program running within base station 24 on microprocessor-assembly 56. However, other implementations, such as a hardware or firmware implementation are also within the scope 25 of the invention.

Data 102 bound for each subscriber station 28 is queued in traffic queues 104 before being sent downstream on BDCH 38 to various subscriber stations 28. Each flow of data 102 can contain a variety of different types of data, such as web pages, FTP data, streaming media, voice over IP data, or other data types as will occur to those of skill in the art. Traffic queues

-11-

104 for subscriber stations 28 are established for each subscriber station 28 that is known to, and is connected to, base station 24 over communication link 32. The example in Figure 5 shows a scenario with four traffic queues 104, each servicing a flow to a corresponding subscriber station 28 (e.g. traffic queue 104a holds traffic bound for subscriber station 28a, etc.) In the example 5 shown in the Figure, traffic queue 104a has five queued packets, 104b has no queued packets, traffic queue 104c has three queued packets, and traffic queue 104d has four queued packets. Traffic queues 104a, 104c, and 104d therefore have backlogged (e.g. – non-zero length queues) flows while traffic queue 104b has no backlogged flow.

Besides holding traffic queues 104<sub>n</sub>, scheduler 100 stores link quality parameter 108<sub>n</sub>, 10 negotiated service share parameter 112<sub>n</sub>, and measured service share parameter 116<sub>n</sub> parameters for each active subscriber station 28<sub>n</sub> with a traffic flow. Furthermore, scheduler 100 stores a fairness factor 120. As described in further detail below, there is at least one instance of fairness factor 120 per sector 36.

The link quality parameter 108<sub>n</sub> holds a suitable measurement of the reception quality 15 experienced at subscriber station 28<sub>n</sub>. ( $\alpha_{\text{subscriber station } 28_n}$ ). In a present embodiment, the value of link quality parameter 108 is the signal to noise (SNR) estimate  $\alpha_{\text{subscriber station } 28} = \frac{E_c}{N_t}$  of at least one suitable channel, where  $E_c$  represents the per chip BDCH channel signal energy at the antenna of subscriber station 28, and  $N_t$  represents the total noise received at antenna 60 of subscriber station 28, the total noise equaling the sum of the average noise density, interference 20 from other interfering cells and sectors, plus multi-path interference. In the present embodiment, each subscriber station 28<sub>n</sub> periodically updates its value for link quality parameter 108 by transmitting its received SNR over an uplink channel such as DDCH 44.

The negotiated service share parameter 112<sub>n</sub> stores the value for an agreed-upon quality of service level ( $\Phi_{\text{neg}}$ ) for subscriber station 28<sub>n</sub>. In the present embodiment, negotiated service share parameter represents a guaranteed data rate (bits/s); however, other definitions of a negotiated service share such as maximum delay before transmitting a waiting packet, or a combination of guaranteed data rate and maximum delay are within the scope of the invention. Subscriber stations 28 with higher values for negotiated service share parameter 112 will receive

-12-

better service than subscriber stations 28 with lower negotiated service share parameter 112. In the present embodiment, negotiated service share parameter 112 is negotiated between base station 24 and each subscriber station 28 when the subscriber station 28 connects to base station 24. However, the means of determining negotiated service share parameter 112 are not particularly limited. For example, negotiated service share parameter 112 can be determined by the service provided based upon the media type about to be transmitted to subscriber station 28, a monthly subscription agreement for subscriber station 28, a fee for service, etc. Other methods of determining negotiated service share parameter 112 will occur to those of skill in the art.

Measured service share parameter 116<sub>n</sub> stores the value of the measured service share (Φ<sub>meas</sub>) for subscriber station 28<sub>n</sub>. The measured service share is the portion of BDCH 38 that has carried packets destined for that particular subscriber station 28<sub>n</sub>. Thus, a larger value for the measured service share parameter 116<sub>n</sub> indicates a higher average data rate delivered to that subscriber station 28<sub>n</sub>. Measured service share parameter 116<sub>n</sub> can be equated to the average bit rate over a fixed interval of time to a particular subscriber station 28<sub>n</sub>.

Finally, fairness factor 120 is an adjustable parameter (F) that represents the scheduling policy and this parameter controls the trade-off between individual flow fairness versus overall throughput over communications link 32. There can be either one instance of fairness factor 120 per sector 36, which is set by a network operator, or there can be one instance of fairness factor 120 per BDCH 38 (in cases where there is more than one BDCH 38 per sector 36). In the current embodiment, F is normalized and ranges from zero to one. A setting of zero indicates a policy that prioritizes the scheduling of data flows 102 to prioritize throughput on the downlink by scheduling data to subscriber stations 28 with the best SNRs, without regard to providing fairness between subscriber stations 28. A setting of one indicating a policy that prioritizes the scheduling of flows 102 to provide fairness between subscriber stations 28s so that all subscriber stations 28 at the same QoS level will receive the same data rate, regardless of their respective SNRs.

Once packets in a traffic queue 104 are scheduled by scheduler 100, they are moved into the blocks 128 of frames 124 of BDCH 38. Normally the spreading factor of a BDCH 38 is predetermined by a network operator and is fixed for every subscriber station 28 that is being

-13-

serviced by a particular BDCH 38. In a current embodiment, a spreading factor of 4 is preferred (thus providing eight blocks 128 per frame 124).

As known to those of skill in the art, the structure of a block 128 can vary according to differences in modulation order, symbol repetition, etc. The number of bits of information carried in each block 128 can vary according to the block structure used. In a current embodiment, each block 128 can carry between three-hundred and twenty and nine-thousand seven-hundred and forty-four bits of information. Generally, blocks 128 with smaller information payloads use lower order modulation and higher symbol repetition to provide greater redundancy for noisy or otherwise poor communication links 32 experienced by subscriber stations 28 with low SNRs. Communication to subscriber stations 28 with better SNRs can employ a block structure that carries more bits of information. In a current embodiment, blocks 128 with different block structures can be carried in the same frame 124.

Referring now to Figure 6, a flowchart for a method of scheduling data transmissions for a BDCH 38 is shown beginning at 200. In the current embodiment, the method described below is run once per scheduled frame 124. However, the frequency of the method is not particularly limited and it can be run more or less frequently than described here, if desired.

At step 200, a scheduling policy is determined by setting a value for fairness factor 120 for sector 36 (between 0 and 1) in base station 24. As described above, a value of 0 indicates a scheduling policy where scheduler 100 prioritizes sector throughput over fairness between subscriber stations 28 and a value of 1 indicates a scheduling policy where scheduler 100 transmits queued packets with maximum fairness to subscriber stations 28.

It is contemplated that, in the majority of deployments, a fairness factor 120 set somewhere between the two bounds of this range (zero and one) will be preferable. For example a setting of 0.5 would provide a policy which provides a reasonable degree of fairness to most subscriber stations 28, while still taking advantage of SNR variances to maximize downlink throughput. It should be noted that even when fairness factor 120 is set to one (indicating maximum fairness), subscriber stations 28 with different QoS levels 112 will be scheduled differently according to their QoS levels, as fairness factor 120 determines fairness in scheduling only with regards to different SNRs. In order to provide totally equal service, QoS levels 112

-14-

should be the same for all subscriber stations 28.

At step 204, scheduler 100 receives the link quality parameter 108 for each subscriber station 28 with backlogged traffic (i.e., any subscriber station 28<sub>n</sub> whose corresponding traffic queue 104<sub>n</sub> has at least one packet in it). Link quality parameter 108 can be updated on each iteration through the method, or can be updated at appropriate longer intervals. At step 208, scheduler 100 calculates a value for the QoS-based priority ( $q_f$ ) for each traffic queue 104 having at least one packet in the queue for this frame.  $q_f$  is a positive number between 0 and 1, with higher values indicating a higher priority for that queue 104. A value for  $q_f$  greater than 0.5 indicates that the data in queue 104 is lagging (i.e., the measured service share parameter 116 is less than the negotiated service share in negotiated service share parameter 112) and a  $q_f$  value less than 0.5 indicates that that particular queue 104 is leading in service (i.e., the measured service share parameter 116 is greater than the negotiated service share in negotiated service share parameter 112). A higher value of  $q_f$  implies that the data in queue 104 requires service sooner.

The value for  $q_f$  is derived by first subtracting a required service share  $\Phi_{freq}$  from the measured service share  $\Phi_{fmeas}$ , dividing the difference by two, then adding 0.5:

$$q_f = \frac{\Phi_{fmeas} - \Phi_{freq}}{2} + 0.5$$

Both  $\Phi_{freq}$  and  $\Phi_{fmeas}$  (described further below) are defined as ranging between zero and one, so  $q_f$  is always a number between zero and one.

The required service share ( $\Phi_{freq}$ ) represents the amount of the service share required for a negotiated service share parameter 112 to be met. For each frame that queue 104 is backlogged, the required service share increases. In the current embodiment  $\Phi_{freq}$  is calculated as follows:

$$\Phi_{freq} = \frac{|w_f| \phi_f}{\sum_{n \in w_f} \phi_f}$$

where the numerator is the product of the number of backlogged frames for flow 102 over a sliding interval of frames ( $w_f$ ) and  $\phi_f$  is the negotiated service share parameter 112 negotiated at setup. The absolute value notation is used to indicate the length of the vector, and  $B(n)$  is the

-15-

indices of the flows which are backlogged during frame  $n$ . The denominator is the sum of negotiated service shares of all backlogged queues 104 over the same sliding interval of frames.

The measured service share ( $\Phi_{fmeas}$ ) is determined by taking all the bits transmitted over BDCH 38 for the selected queue 104 during a sliding window interval, and dividing it by all the bits transmitted over BDCH 38 during the same sliding window interval for every queue 104. In the current embodiment,  $\Phi_{fmeas}$  is calculated as follows:

$$\Phi_{fmeas} = \frac{\sum_{n \in W} b_f(n)}{\sum_{n \in W} \sum_{r \in B(n)} b_r(n)}$$

where  $b_f(n)$  is the number of bits transmitted from flow  $f$  during frame  $n$ .

At step 212, scheduler 100 computes the throughput priority ( $t_f$ ) for each backlogged queue 104 for this frame.  $t_f$  is a normalized interpretation of the SNR for a particular flow 102 ( $f$ ), yielding a value between 0 and 1, where 0 indicates a minimum SNR ratio and 1 indicates the maximum possible SNR. Mathematically,

$$t_f = \frac{b(\alpha_{ss(f)}) - b(\alpha_{\min})}{b(\alpha_{\max}) - b(\alpha_{\min})}$$

where  $b(x)$  is the number of bits in block 128 at SNR  $x$ , as determined by the block structure.

At step 216, scheduler 100 computes the total priority function ( $p_f$ ) for each backlogged queue 104 by calculating a value derived from adding the QoS priority ( $q_f$ , as determined at step 208) to the throughput value ( $t_f$ , as determined at step 212), where each priority is multiplied by fairness factor 120 or an inverse of the fairness factor 120 respectively.

In the current embodiment,  $p_f$  is calculated as follows:

$$p_f = Fq_f + t_f(1-F), f \in B$$

where  $B$  is the set of all backlogged queues 104.

In the above equation,  $F$  is the fairness factor 120. When fairness factor 120 is zero, then the product of  $F$  and  $q_f$  is zero and the term ( $\alpha'_{ss(f)}(1-F)$ ), which is the throughput value, is maximal. When fairness factor 120 is one, then  $q_f$  is maximal and the throughput value is multiplied to a product of zero.

At step 220, all the backlogged queues 104 are sorted by the total priority function ( $p_f$ )

-16-

determined in step 212 in descending order,  $\{p'\} | p_{p'(i)} \geq p_{p'(j)}, i < j$ .

At step 224, starting with the queue 104 with the highest total priority value  $p'(1)$ , scheduler 100 computes the maximum number of blocks 128, referred to herein as  $(m'(i))$ , that will be allocated to this queue 104 by calculating a percentage of blocks 128 available to it,

5 based upon the priority of this queue 104 relative to the sum of all total priority values for all backlogged queues. More specifically, scheduler 100 multiplies the number of blocks 128 in frame 124 ( $M$ ) by the priority function ( $p_f$ ) for this queue 104 (as determined in step 212), then dividing the result by the sum of all priority functions (also determined in step 212) for all backlogged queues. In the current embodiment the maximum number of blocks 128  $m'(i)$  is  
10 determined as follows:

$$m'(i) = \text{round} \left( \frac{Mp'(i)}{\sum_{f \in B} p'(f)} \right)$$

where the round() operator rounds the number of scheduled blocks,  $m'(i)$ , to the closest integer.

At step 228, starting with the highest priority queue 104 ( $p'(1)$ ), scheduler 100 allocates up to  $m'(1)$  blocks 128. Fewer blocks 128 are allocated if the number of bits available for 15 scheduling to queue 104 ( $p'(1)$ ) requires fewer than  $m'(1)$  blocks 128. Step 220 is repeated for all backlogged queues 104 ( $p'(2)$ ),  $p'(3)$ ), etc., until either all backlogged queues 104 are cleared or all the blocks 128 in frame 124 are scheduled. Once step 220 is completed, the method returns to step 204 to schedule the next frame 124. The method continues whenever data is present to be scheduled for transmission.

20 While the embodiments discussed herein are directed specific implementations of the invention, it will be understood that combinations, sub-sets and variations of the embodiments are within the scope of the invention. For example, in the current embodiment, only a single queue is scheduled for each traffic block 128. However, multiple queues can be scheduled in a traffic block 128 and are within the scope of the invention. If multiple subscriber stations 28 were scheduled in the same block 128, the block structure must be suitable to satisfy the SNR 25 requirements of all subscriber stations 28 scheduled. One way to ensure this, for example, is to only allocate subscribers that have an SNR that is greater than the first subscriber that is

allocated to the block.

It is contemplated that scheduler 100 may run less frequently when the total amount of traffic in all queues 104 is below a certain threshold (representing a low amount of data traffic volume), as to reduce latency for microprocessor-router assembly 36. Alternatively, when the 5 total amount of traffic is below a certain threshold, scheduler 100 may elect to suspend its scheduling method (as described with regards to Figure 6), and instead schedule traffic according to some other scheduling method such as FIFO, in order to reduce latency for each queue 104

It is further contemplated that scheduler 100 can run more frequently than once per frame 10 124, where the scheduling frequency is a multiple integer of the duration of block 128. This will permit the scheduler to perform well in situations where the coherence time of the channel is lower, such as when the carrier frequency is higher or if there is a greater amount of motion of pedestrians or other traffic near to the subscriber or the base station.

It is also not intended that the present invention be limited to use with the particular radio 15 based system described above, nor to radio-based systems in general and it is believed that the present invention can be advantageously employed with any system for scheduling data for transmission from a single node to one or more of a plurality of other nodes over a shared communications link. Use with optical, wireline and other shared communication links is contemplated.

20 The above-described embodiments of the invention are intended to be examples of the present invention and alterations and modifications may be effected thereto, by those of skill in the art, without departing from the scope of the invention which is defined solely by the claims appended hereto.